When Maximizing the Test Becomes Paramount, Does Learning Become Secondary?

A Case For Assessment/Evaluation of Programs vs. Testing of Students

DR. ANTHONY A. SCAFATI

> I never find grading a satisfactory experience. It deals with teaching, evaluation, accreditation, indoctrination, control, and unthought. It's demeaning to all parties. I get ulcers on the inside of my bottom lip every time I do it.
>
> *—Dr. Jerry Harvey(1980) Sermon #13*

Jerry Harvey is a funny guy, but there is much wisdom in his words. Most of us who read the *Program Manager* are graduates of some course at DSMC or one of the consortium schools. Many of us are responsible for teaching our juniors or contemporaries the tricks of the trade. We are products ourselves, of a fine American education system that in spite of some of the criticism, produces graduates that continue to compete and win in the world in all disciplines. However most institutions of learning were not established to meet the needs of a unique student body, from a well-defined organization, and a known work environment. They were established to teach "individuals" who had broad needs, and would work and function in an unaccountable variety of workplaces. The goal then, for American education, was to prepare individuals for success in the workforce.

Best and Brightest?

Since the focus was on individual success, it was a natural next step for American education to evaluate individuals against their peers, thus the lionization of grades/awards/class standings, Valedictorians etc. These so called motivators became the norm and have served well as an easy way to determine who are the so called "best & brightest."

Before we can accept that the grading systems in our schools actually determine who are the "best and the brightest," we have to ask ourselves: "Measured against what?" In contrast, the consortium schools have defined customers with very well-defined

WE NEED TO DEVELOP A PROCESS THAT PROVIDES "DASH BOARD" INDICATORS THAT LEARNING IS HAPPENING. WHILE THE LESSON IS IN PROGRESS, WE CAN GAUGE WHETHER THE STUDENTS AS A GROUP HAVE LEARNED THE SUBJECT AT THE DESIRED LEVEL (FORMATIVE ASSESSMENT).

Scafati is a Faculty Department Chair, Education Department, Faculty Division, DSMC.

needs, who return to work environments of a specific and known nature. Therefore our curriculum is designed (or should be) to meet the specific needs of the customers and the environment in which they work.

This environment relies on teams of diverse people with varying fields of expertise having a shared vision. Assessment of learning in this environment should be competency (performance) based on specific behaviors required in that work environment.

To be sure, American Higher Education is examining assessment/evaluation versus testing, and there is a wealth of research in this area. Some of that research I have used in writing this article. But the purpose of this article is not to change American Higher Education, but to enlighten individuals and institutions on the benefits of assessing/evaluating processes (curricula) versus testing individuals in general, and specifically the consortium schools. To do that, let me first tell you a true story about how it was in the old Program Management Course (PMC) and how it is now in the Advanced Program Management Course (APMC) at DSMC.

Currently, [testing] it is used to assess the learning that is or is not taking place while the class is in process (formative). If the instructors determine that the learning has not taken effect to the degree and level desired, then they have the opportunity to "teach" on the spot.

In the Beginning — PMC

The PMC course was a fine course, which brought together military and civilian students from five Services and industry, with diverse backgrounds, experience, and education. Admittedly, PMC had a challenging curriculum covering functional areas from Management Development to Systems Engineering. The faculty and staff worked diligently to ensure that the subject matter was current with the best practices of the industry and the latest policy.

A class at DSMC was divided into 30-student sections, each section into 5 or 6 student work groups. During the conduct of the course, sections and work groups were divided and then reconstituted in order to provide each individual student an opportunity to be exposed to as many other students and their unique experiences and perspectives as possible.

The curriculum mirrored how we did work in the program office. The students had opportunities to perform in experiential exercises that simulated a real work environment. Some of these exercises took as long as 75 hours of class time, had a multi-objective approach, and required changing of student leadership, and roles.

Our Management Development course emphasized teams, optimum team functioning, the strength of a team, the importance of a team in acquisition management, and one's responsibilities as a team member. The entire purpose of the structure of the section/work groups, the design of the curriculum and the exercises, and other experiential exposure was to prepare graduates to be "competent" members of the Acquisition Corps. We wanted DSMC students to leave with more than mere knowledge and understanding of weapons systems acquisition, but to also apply, analyze, synthesize, and evaluate programs while a member of a team.

But Are We Teaching Them Anything?

Of course we all know that being a member of a team may mean we must sublimate personal goals for group goals. It is an attitude, a value, that puts the mission ahead of the individual. We were doing well, with high satisfaction ratings from all customers. Then someone asked the question, "How do we know we are teaching anyone anything?"

This is a legitimate question that should be asked by any organization in the education business. However since we all come from an education system that emphasizes and rewards individual accomplishments rather than group successes, we turned to "testing" as the way we were going to measure the learning effectiveness of this institution. Was testing effective?

Did we gain insight into the reliability and validity of our educational processes? Did we foster team membership? Did we create a learning environment?

What we created was a fiercely competitive atmosphere where students at times withheld information from other students; "gamed" the test by studying old tests to determine which functional area had the most questions; how frequently that question appeared on the test to determine the probability of being on the next; and worst of all, in my opinion, they traded off learning of the subject matter to learning the test.

It was commonplace, even encouraged, for sections to divide the Desired Learning Outcomes (DLO) among the section "experts." Those individuals would write a short answer to address a particular DLO so that the majority of the section could learn an "abridged" answer which would satisfy a question on the test. No one wanted, or in their mind could afford career-wise, a low score.

Is Testing Paramount, And Learning Secondary?

In the end, maximizing the test became paramount; learning became secondary. We tried many ways to reduce what we labeled "test anxiety." We eliminated the grades and replaced them with "pass" or not yet. We attempted to eliminate the "finals" atmosphere of the end-of-course test by giving smaller and more frequent tests. We went from predominantly "multiple choice/fill in the blank" tests to more comprehensive essay types. In the end, we abandoned testing students to assessment of the curriculum process.

That is not to say that there is no place for formative (in-process testing) or summative (end-of-course assessment), and that there is no testing in the new PMC (now called the Advanced Program Management Course, or APMC). There is room for some of both to be sure. Currently, it is used to assess the learning that is or is not taking place while the class is in process (formative). If the instructors determine that the learning has not taken effect to the degree and level desired, then they have the opportunity to "teach" on the spot.

Test instruments are also used by some functional areas at the end of the course of instruction to assess whether the desired learning outcomes have taken place and to what degree (summative). If, across the entire class, the results of the assessment indicate that the desired learning has not taken place to the degree desired, it may mean there is a design flaw. Obviously, that should be corrected prior to the next iteration of the subject matter.

It is important to understand that it is not necessarily the evaluation "instrument" that is objectionable, but the use of the data collected by that instrument. If the instrument is used to assess the progress of learning of the group (aggregate of individuals' performances) in order to make immediate adjustment to the process, or to determine the validity and reliability of the design, it is worthwhile. If the instrument and the resultant data are used to seek the "best and the brightest," it becomes an incentive to do well on the test and to abandon adult learning. This is an example of what I believe is Maslow's paradigm of the "Self-Actualization" (Combs, Avila, Purkey, 1971) in that if we are to be hired/fired/promoted/demoted based on a "test," survival takes precedent over self-actualization, and our priority will be to maximize the test, not learn all we can.

Testing — A Continual Struggle

In the new APMC the designers have abandoned "testing" individuals for the purpose of determining the best and the brightest and embraced assessment and evaluation. We have adopted a quality approach to education. That is, if students are competent, and have the appropriate background in education and experience, and the curriculum (process) is sound, then they will learn to the level required, and we do not have to test each individual.

Now don't take my word for it; there is and has been a continual struggle within the education community as to the value and purpose of testing. This article is written to address this issue as it pertains to education in general and, specifically, to the mission and purpose of teaching acquisition management to adults.

To most of us who have been brought up in an educational system that tends to *test* the progress of the student rather than effectiveness of the program, we intuitively assume that traditional grading accurately displays the quality and the quantity of learning that has taken place. Contemporary research, however, clearly refutes this premise (Knowles, 1980; Nadler, 1982; Harvey, 1977).[1,2,3] Traditional grading is weak enough with adolescents; it becomes less relevant with adults (Knowles, 1980; Nadler, 1982).

What the research seems to suggest is whether we want to evaluate or not evaluate, who we evaluate, and how we evaluate is really a conflict of values (Knowles, 1980, p. 201). On one hand, we have the behaviorists who need hard data, proof, science, and who value control. On the other hand are those who espouse Maslow's values of self-actualization, free play of natural forces, and place a high value on the humanistic aspects of management. Since both of these points of view are present and valued in our society, a position in the middle is probably what will take us the furthest in evaluation as adult educators (Knowles, 1980).

Two Dominant Themes Emerge — Quantification and Involvement

Experienced faculty observed this phenomenon in the PMC course when we had three major tests. As I previously mentioned in this article, we observed "gaming" of the test by students in a section who divide the DLOs amongst themselves. Therefore, one out of 30

students might know the DLO well; the other 29 might know only enough to answer that question on the test. Additionally, the class would gather former tests and study the design. They would ascertain how many times a question showed up on a final to determine the probability that the question would be on the next final. The resultant energy expended and stress generated actually detracted from the learning opportunity. This is the antithesis of an adult learning atmosphere.

Knowles (1980) goes on to say that how much and what type of evaluation you will apply to adult education is simply but unequivocally, a product of our philosophy and definition of education. Therefore, if instructors define their responsibilities as "[for] making changes in a human being," then they do incur an obligation to efficiently obtain data to ensure they are producing maximum change, in the shortest amount of time, for the least cost. The dominant theme in this case would be *quantification*.

If, however, one's definition of adult education is facilitating and providing resources for *self-directed* inquiry and self-development, one incurs an obligation to involve the students in collecting the data that will enable them to assess the effectiveness of the program in helping them meet their objectives. The dominant theme in that case is *involvement*. The difference is simply the conflict of pedagogy versus andragogy (Knowles, 1980).

Cremin (1976, pp. 88-89)[4] speaks eloquently on the heart of the problem when he admonishes us to develop better techniques for monitoring and assessing education. He states:

> For all our sophistication in *testing* [emphasis applied] — and we have made tremendous strides in the last decade or so — our instruments are still imprecise about what should be evaluated and to what purpose. They deal almost exclusively with the cognitive aspects of learning. They tend to *separate individuals for the purposes of selection* [emphasis applied] rather than providing information on the performance of the education system as a whole...

...if we are to be hired/fired/promoted/demoted based on a "test," survival takes precedent over self-actualization, and our priority will be to maximize the test, not learn all we can.

Whitlock (1986, pp. 74-76)[5] writes in a fascinating little book, *Educational Myths I Have Known and Loved*, a significant chapter entitled, "The Myth That Grades Are Important." In this chapter, he points out that "grades are always relative." At the present time, and as long as grade inflation remains with us, even the relative value of grades has disappeared. He further argues that:

> Faculty members depend on grades almost as much as students do — perhaps even more. They are a crutch, and a traditional crutch on which there is enough agreement to make the system work, even though it may be ultimately dishonest. (For example, why are an 89 and an 80 the same grade when 80 and 79 are different?)

"The Art of Gaming"

Whitlock's point of view corresponds with Dixon (1990, p. 32)[6] who intimates that when instructors are rewarded for high ratings [could be high grades], they tend to modify their behavior to ensure student enjoyment. I have observed professors here and in other institutions who teach to the test, not to the subject. Why, haven't we institutionalized the Scholastic Aptitude Test (SAT)? Haven't we institutionalized it to such a degree that parents pay exorbitant amounts of money to prepare their children to do well on the test? We have whole industries who prepare people for the SAT, LSAT, GMAT, GRE, etc. My own research has shown that as a predictor of success in college, the SAT is only the third best, preceded by Grade Point Average, and of all things, "Family Income." Additionally, the verbal scores in the SAT were the predictor, and in the population I studied the math portion predictability was not significant. (Scafati, 1990).[7]

Continuing with Whitlock's myths (1986, p. 75), he contends that some of the best students receive "C's." The reason, he insists, is because "they refuse to 'learn the teacher' rather than the subject." Another name for this process could be "The Art of Gaming," which is a euphemism for learning how to play the teacher rather than learn the material.

True Definition of Evaluation

We have come a long way since 1979 in evaluation processes. After reviewing the literature, it has become abundantly clear to this author that the true definition of evaluation is dependent on the purpose of the evaluation. The purposes are many and, therefore, the definitions are varied. In the following paragraphs, I have tried to describe some of the current thinking regarding evaluation in training and education.

Knowles, in his 1980 book, *The Modern Practice of Adult Education*," quotes Stufflebeam in that evaluation serves two purposes:

- The first is *accountability* — justification of the value of the program to employers, sponsors, the clientele, or society. This he calls summative evaluation.
- The second purpose is to improve *decision making* by providing information to the [course] program managers that will enable them to improve the quality of the program. This calls for formative evaluation.

Both types of evaluation must take into account the four elements of the program: goals, design, process, and product.

Nadler (1982) emphasizes that the purpose of evaluation is to ensure the design is valid and reliable and that modification to improve outcomes is present. At no time does he advance the proposition that evaluation is to determine the standing of the student or the competency of the instructor.

Clark (1989)[8] states that the purpose of assessment is to determine the effect of the training. At no time does she allude that we need to test individuals except as a source of data to determine the effectiveness of the experience.

Phillips (1983), in the *Handbook of Training Evaluation and Measurement Methods*,[9] lists eight purposes — most of which are in consonance with the literature: (He speaks of "HRD" [Human Resource Development], which is in essence *training*.)

- To determine whether the program is accomplishing its objectives. *(Reliability)*
- To identify the strengths and weaknesses in the HRD process.
- To determine the cost/benefit ratio of an HRD program.
- To decide who should participate in future programs.
- To identify which participants benefited most from the program.
- To reinforce major points made to the participant.
- To gather data to assist in marketing future programs.
- To determine if the program was appropriate. *(Validity)*

What is clear in all these purposes is that nowhere do we see that the purpose is for selection or discrimination among or between students. In general, the purpose is to assess the reliability and the validity of the program and to provide feedback to the student and the organization on the degree and quality of the learning that has taken place. Finally, the assessment process is another opportunity (in an experiential way), of reinforcing learning or moving up the taxonomy of learning.

Dixon (1990, p. 27) cites Kirkpatrick et al in that the purpose of evaluation remains the same: to improve the learning experience and not to measure, assess, or grade the student.

Assumptions and other myths about assessment and evaluation: Who Should Evaluate? Every person who is involved in any way with the development or the execution of an adult program should evaluate the program from their personal perspective (Knowles, 1980, pp. 204-205).

As I See It

In my journey through the literature referenced in this article, the following conclusions have become clear:

- Evaluation is valuable.
- There are ethical considerations, in that an institution of learning has an ethical responsibility to provide the product it contracts to deliver.
- The only way to ensure this is being done in a quality way is to evaluate the process and the outcomes.
- The purpose of evaluation is not to define the difference among or between students.
- The purpose of evaluation is to ensure reliability and validity of the program.
- Both formative and summative evaluation are important.
- The real proof of the program is measured after some time has elapsed, and where the job for which the program was designed, is performed.

Recommendations

From the above conclusions, the following recommendations are made for all institutions whose purpose is to graduate persons who can effectively "perform" in the workforce.

- That they develop an evaluation program with the express purpose of improving the product provided to their customers. (The customers in this case are the students, their immediate supervisors, and the acquisition community.)
- That the types of evaluation conducted would be Formative, Summative, and performance-based.
- That the evaluation process be concurrent with the design of the courses, and indeed, the design process be iterative so that any changes to form or process will be incorporated during course upgrade.

My Conclusions

I believe we at DSMC are heading in the right direction. We still need to do more work in assessing our process of creating an adult learning environment. We need to develop a process that provides "Dash Board" indicators that learning is happening. While the lesson is in progress, we can gauge whether the students as a group have learned the subject at the desired level (formative assessment). We must also develop a process that determines whether the objectives of the course have been reached by the class to the desired learning level (summative assessment), and be prepared to change the design when it does not meet the outcomes reliably. If we accomplish these goals, we will not only be following some of the best minds in education, but also adhering to the teachings of some of the best minds in quality.

In conclusion, if the material we use to produce something is sound (the student); the process we use to produce the product is sound (the curriculum); the equipment is appropriate for the task (course materials); and the worker has the required skills (faculty), then there is no need to inspect (test) at the end of production (graduation)! We will have gathered enough empirical data along the way to assess the learning, improve the process, and satisfy those who need proof of the results.

REFERENCES

1. Knowles, Malcolm S., *The Modern Practice of Adult Education*, Chicago: Follett Publishing Company, 1980.

2. Nadler, Leonard, *Designing Training Programs, The Critical Events Model*, Reading, Massachusetts: Addison-Wesley Publishing Company, 1982.

3. Harvey, Jerry B., Learning to Not Teach," *Exchange: The Organizational Behavior Teaching Journal*, Spring 1977, pp. 19-21.

4. Cremin, Lawrence A., *Public Education*, New York: Basic Books, Inc., Publishers, 1976.

5. Whitlock, Baird W., *Educational Myths I Have Known and Loved*, New York: Schocken Books, 1986.

6. Dixon, Nancy, "Evaluation: A Tool for Improving HRD Quality," University Associates Inc., 1990, pp. 27-37.

7. Scafati, A., "A Study to Determine Selection Criteria for Predicting Individual Success in the Navy and Marine Corps BOOST and Officer Selection Programs," Ed.D. Dissertation, 1990.

8. Clark, Ruth Colvin, *Developing Technical Training*, Reading, Massachusetts: Addison-Wesley Publishing Company, 1989.

9. Phillips, Jack, *Handbook of Training and Evaluation and Measurement Methods*, Houston, Texas: Gulf Publishing Company, 1983.

Photo by Richard Mattox